Function-Valued Reproducing Kernel Hilbert Spaces and Applications

Hachem Kadri
SequeL Project
INRIA Lille - Nord Europe
Villeneuve dAscq, France
hachem.kadri@inria.fr

Philippe Preux
Sequel Project/LIFL
INRIA Lille/Université de Lille
Villeneuve dAscq, France
philippe.preux@inria.fr

Emmanuel Duflos
Sequel Project/LAGIS
INRIA Lille/Ecole Centrale de Lille
Villeneuve dAscq, France
emmanuel.duflos@ec-lille.fr

Stéphane Canu
LITIS
INSA de Rouen
St Etienne du Rouvray, France
scanu@insa-rouen.fr

Manuel Davy
LAGIS/Vekia SAS
Ecole Centrale de Lille
Villeneuve dAscq, France
mdavy@vekia.fr

Hilbert spaces of scalar functions with reproducing kernels were introduced and studied in [2]. Due to their crucial role in designing kernel-based learning methods successfully applied in several machine learning applications, these spaces have received considerable attention over the last two decades [14]. More recently, interest has grown in exploring Hilbert spaces of vector random functions for learning vector-valued functions [9, 4], even though the idea of extending Reproducing Kernel Hilbert Spaces (RKHS) theory from scalar to vector-valued functions is not new and dates back at least to [15]. In [5] Hilbert spaces of vector-valued functions [9] are described and matrix-valued reproducing kernels [8] are constructed to learn many related regression or classification tasks simultaneously [1]. Similarly, there are also some works where approaches based on this framework have been proposed to estimate a two or three-dimensional vector-valued function [6]. In this work, we propose to address the general case where the output space is a space of vectors with large dimension, even infinite. In this latter case, the output space is a space of functions and data which belong to that space are usually called functional data, so we focus here on extending RKHS theory from vector to functional data. More precisely, we focus on reproducing kernel Hilbert spaces whose elements are function-valued functions and we demonstrate how basic properties of real RKHS can be restated in the functional case. The framework developed is valid for any type of input data (vector or function). However, we are interested is the case where both input and output data are functions. In this setting, we consider the problem of estimating a function f such that $f(x_i) = y_i$ when observed data $(x_i, y_i)_{i=1,\ldots,n}$ are assumed to be elements of infinite dimensional Hilbert spaces. In the following we denote by $\mathcal{G}_x$, and $\mathcal{G}_y$ the domains of x_i and y_i respectively. $X = \{x_1, \ldots, x_n\}$ denotes the training set with corresponding targets $Y = \{y_1, \ldots, y_n\}$. Since $\mathcal{G}_x$ and $\mathcal{G}_y$ are spaces of functions, the problem can be thought of as an operator estimation problem, where the desired operator maps a Hilbert space of factors to a Hilbert space of targets. We can define the regularized operator estimate of $f \in \mathcal{F}$

$$f_\lambda \triangleq \arg\min_{f\in\mathcal{F}} \sum_{i=1}^{n} \|y_i - f(x_i)\|^2_{\mathcal{G}_y} + \lambda\|f\|^2_{\mathcal{F}}$$

In the first part of this work, we provide a solution to this minimization problem in a reproducing kernel Hilbert space $\mathcal{F}$ of function-valued functions on some infinite-dimensional input space $\mathcal{G}_x$. We start by introducing function-valued reproducing kernel Hilbert spaces and their correspondence to positive operator-valued kernels. Then we discuss the construction of operator-valued kernels. These kernels which take two functions in $\mathcal{G}_x$ as arguments and output an operator in $\mathcal{L}(\mathcal{G}_y)$, where $\mathcal{L}(\mathcal{G}_y)$ is the set of bounded linear operators from $\mathcal{G}_x$ to $\mathcal{G}_x$, can be built as follows. First one can attempt to build an operator $T^h \in \mathcal{L}(\mathcal{G}_y)$ from a function $h \in \mathcal{G}_x$ [3]. We call h the characteristic function of the operator T^h [13]. In this first step, we are building a function $f : \mathcal{G}_x \longrightarrow \mathcal{L}(\mathcal{G}_y)$. The second step can be performed in one of two ways. Either we build h from a combination of two functions h_1 and h_2 in $\mathcal{H}$, or we combine two operators created in the first step using the two characteristic functions h_1 and h_2. The second way is more difficult because it requires the use of a function which operates on operator variables. To build an operator-valued kernel and then constructing a function-valued reproducing kernel Hilbert space, the operator T^h is of paramount importance. Choosing T presents two major difficulties. Computing the adjoint operator is not always easy to do, and then, not all operators verify the Hermitian condition of the kernel. The kernel must be nonnegative, this property is given according to the choice of the function h. Finally, we provide examples of operator-valued kernels and we show their use for solving the minimal norm interpolation problem of function-valued functions within the RKHS framework developed.

The second part of this work is devoted to describe some applications where the framework developed for learning function-valued functions can be of practical interest. Observing and saving complete functions as a result of random experiments is nowadays possible by the development of real-time measurement instruments and data storage resources. Functional data arise in a number of scientific fields associated with continuous-time processes whose final outputs are samples of (discretized) functions. It is true that the measurement process itself very often provides a vector rather than a function, but the vector is really a discretization of a real attribute, which is a function. Furthermore, if the discretization step is small, the vectors may become very large. We think that handling these data as what they really are, that is functions, is an interesting path to investigate. Methods working with discretized data more or less ignore their functional character, and thus extending learning methods from multivariate to functional data could lead to further progress in several practical problems of machine learning and applied statistics.

Typical areas of research where one has to deal with functional data are meteorology, physiology and speech processing. A class of problems in these fields which illustrate the potential of adopting a function-valued RKHS approach is functional regression [12] when data attributes as responses are functions. In this setting, an example is a couple $(x_i(s), y_i(t))$ in which both $x_i(s)$, and $y_i(t)$ are real functions, that is $x_i(s) \in \mathcal{G}_x$, and $y_i(t) \in \mathcal{G}_y$ where $\mathcal{G}_x$, and $\mathcal{G}_y$ are real Hilbert spaces. We notice that s and t can belong to different sets. This setting naturally appears when we wish to predict the evolution of a certain quantity in relation to some other quantities measured along time. Most previous works on this problem suppose that the relation between functional responses and predictors is linear. In this case, the functional regression model is an extension of the multivariate linear regression model

$$y(t) = \alpha(t) + \beta(t)x(t) + \epsilon(t)$$

for a regression parameter function β. In this model, known as the concurrent model, the response y and the covariate x are both functions of the same argument t, and the influence of a covariate on the response is concurrent or point-wise in the sense that x only influences $y(t)$ through its value $x(t)$ at time t (see chapter 14 of [12]). An extended linear model in which the influence of a covariate x can involve a range of argument values $x(s)$ takes the form (see chapter 16 of [12])

$$y(t) = \alpha(t) + \int x(s)\beta(s,t)ds + \epsilon(t)$$

In this setting, an extension to nonlinear contexts can be achieved using Hilbert spaces of function-valued functions and functional reproducing kernels [7]. We discuss examples, related to the fields mentioned above, where we feel there is practical need for this extension. In meteorology, there is a continuing interest in analyzing and understanding relationships between climate data. A first example deals with weather data. For 35 weather stations of Canada, the daily temperature and precipitation were averaged over a period of 30 years. The goal is to predict the complete log daily precipitation profile of a weather station from information on the complete daily temperature curves [12]. The second example consists in studying the dependence of the acceleration of the lower lip in speech on neural activity, as measured by electromyographical (EMG) recording. This kind of study is of particular relevance for physiologists who aim to understand how the brain controls movement. Input-output data consist in 32 records of the movement of the center of the lower lip when a subject was repeatedly required to say the syllable “bob”, embedded in the phrase, “Say bob again” and the corresponding EMG activities of the primary muscle depressing the lower lip, the depressor labii inferior (DLI) [11]. The goal is to predict lip acceleration from EMG activities. As third and a final example we consider the problem of learning the acoustic-to-articulatory mapping, also known as speech inversion [10, 16]. It is a speech processing related problem that has attracted the attention of several researchers. The goal is to investigate ways to exploit articulatory informations, which are described by electromagnetic articulography (EMG) trajectories or vocal tract time functions, to improve both speech technology and understanding. A successful solution could find numerous applications, such as helping individuals with speech and hearing disorders and improving speech recognition systems.

References

[1] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3):243–272, 2008.

[2] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

[3] S. Canu, X. Maru, and A. Rakotomamonjy. Functional learning through kernel. *in Advances in Learning Theory: Methods, Models and Applications. NATO Science Series III: Computer and Systems Sciences*, 2003.

[4] C. Carmeli, E. De Vito, and A. Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem. *Analysis and Applications*, 4:377–408, 2006.

[5] T. Evgeniou, C. A. Micchelli, and M. Pontil. Learning multiple tasks with kernel methods. *Journal of Machine Learning Research*, 6:615–637, 2005.

[6] M. Ha Quang, S. H. Kang, and T. M. Le. Image and video colorization using vector-valued reproducing kernel Hilbert spaces. *Journal of Mathematical Imaging and Vision*, 37(1):49–65, 2010.

[7] H. Kadri, E. Duflos, P. Preux, S. Canu, and M. Davy. Nonlinear functional regression: a functional rkhs approach. In *Y.W. Teh and M. Titterington (Eds.), Proceedings of The Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS) 2010, JMLR: W&CP 9*, pages 111–125, Chia Laguna, Sardinia, Italy, 2010.

[8] C. A. Micchelli and M. Pontil. Kernels for multi-task learning. In *Advances in Neural Information Processing Systems 17 (NIPS 2005)*, pages 921–928, 2005.

[9] C. A. Micchelli and M. Pontil. On learning vector-valued functions. *Neural Computation*, 17:177–204, 2005.

[10] V. Mitra, Y. Ozbek, H. Nam, X. Zhou, and C. Y. Espy-Wilson. From acoustics to vocal tract time functions. In *ICASSP ’09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4497–4500, Washington, DC, USA, 2009.

[11] J. O. Ramsay and B. W. Silverman. *Applied Functional Data Analysis*. Springer Verlag, New York, 2002.

[12] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis, 2nd ed*. Springer Verlag, New York, 2005.

[13] W. Rudin. *Functional Analysis*. McGraw-Hill Science, 1991.

[14] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, MA, USA, 2002.

[15] L. Schwartz. Sous-espaces hilbertiens d’espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d’Analyse Mathématique*, 13:115–256, 1964.

[16] A. Toutios and K. Margaritis. Learning articulation from cepstral coefficients. In *10th International Speech and Computer Conference (SPECOM’05)*, Patras, Greece, 2005.